SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I5P129

Volume 12 Issue 5, 350-362, May 2025

Original Article

Automated Detection of Lower Back Pain Using Machine Learning and SMOTE-Based Data Augmentation

P. Praveen[1], Mallikarjunaswamy M.S.[1*], Mahabaleshwar Mamadapur[2]

[1]*Department of Electronics & Instrumentation, Sri Jayachamarajendra College of Engineering, JSS Science and Technology University, Mysuru, India.*
[2]*Department of Clinical Immunology and Rheumatology, JSS Medical College and Hospital, JSS Academy of Higher Education and Research, Mysuru, India.*

**Corresponding Author : msm@sjce.ac.in*

Received: 19 March 2025 Revised: 20 April 2025 Accepted: 22 May 2025 Published: 27 May 2025

Abstract *- Low Back Pain (LBP) is a leading global health concern, affecting up to 80% of individuals at some point in their lives and ranking among the most common causes of chronic disability and work absenteeism. Despite advancements in treatment, accurate and scalable diagnostic tools remain limited. Traditional diagnostic methods rely heavily on clinical expertise and imaging, which are often time-consuming, subjective, and inaccessible in resource-limited settings. Recent literature underscores the potential of Machine Learning (ML) for automating LBP detection, but challenges such as imbalanced datasets and insufficient model generalizability persist. This study introduces a robust ML pipeline for automatic LBP classification using data from the publicly available international dataset - Kaggle. The workflow incorporates data type normalization, outlier elimination, and feature distribution analysis, followed by class rebalancing through the Synthetic Minority Oversampling Technique (SMOTE). Three ML classifiers-Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN)-are trained and evaluated on both imbalanced and SMOTE-balanced datasets. Experimental results demonstrate a significant boost in classification performance post-balancing, with the ANN model achieving the highest accuracy (96.43%) and F1-score (96.47%). This work confirms that integrating effective preprocessing with optimized model selection can deliver accurate, scalable, and automated LBP detection-offering a meaningful step toward smarter musculoskeletal diagnostics.*

Keywords *- Artificial Neural Network, Imbalanced data, Lower Back Pain, Machine Learning, SMOTE, Outliers.*

1. Introduction

Low Back Pain (LBP) is a prevalent musculoskeletal condition that affects a wide range of age groups, significantly impairing quality of life and reducing workplace productivity. It is among the most frequent reasons for primary care visits. While some cases can be attributed to identifiable causes such as trauma, infection, or spinal anomalies, the majority are nonspecific with no clearly defined aetiology. The aetiology of low back pain is summarized in Table 1. The 2010 Global Burden of Disease Study ranked LBP among the top ten diseases and injuries contributing to global Disability-Adjusted Life Years (DALYs) [1]. Data from the National Center for Health Statistics reveal that 27% of adults report experiencing LBP, compared to 15% reporting neck pain or severe headaches/migraines [2]. The condition often radiates to adjacent regions, including the thighs, hips, and lower limbs, further exacerbating discomfort. Notably, up to 80% of individuals may experience LBP during their lifetime, with degenerative changes in the lumbar spine, such as intervertebral disc degeneration, being a primary contributor [3]. Imaging modalities like X-ray, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans are conventionally employed to assess structural abnormalities of the spine [4]. Recently, there has been a growing interest in utilizing Machine Learning (ML) techniques to augment diagnostic accuracy and automate the classification of normal versus pathological conditions. ML algorithms, when trained on medical imaging data including MRI, X-rays, and even thermal images, can identify subtle patterns and abnormalities not easily perceptible by human experts [5]. Integrating ML into LBP diagnostics has the potential to not only improve diagnostic efficiency but also facilitate early detection during routine health check-ups. This is particularly promising for asymptomatic individuals or those with non-specific symptoms, where early intervention could mitigate long-term disability and healthcare burden. Therefore, the implementation of ML-based diagnostic tools could revolutionize the current approach to LBP assessment by enabling timely, cost-effective, and objective screening [6].

Table 1. Differential diagnosis of lower back pain: etiological categories [7]

Category	Subcategories
Mechanical	Lumbar spondylosis, Disk herniation, Spondylolisthesis, Spinal stenosis, Fractures (mostly osteoporotic), Nonspecific (idiopathic)
Neoplastic	Primary, Metastatic
Inflammatory	Spondyloarthritis
Infectious	Vertebral osteomyelitis, Epidural abscess, Septic diskitis, Herpes zoster
Metabolic	Osteoporotic compression fractures, Paget's disease
Referred Pain to Spine	From major viscera, retroperitoneal structures, urogenital system, aorta, or hip

2. Literature Survey

Low Back Pain (LBP) continues to be one of the most pervasive musculoskeletal disorders worldwide, affecting up to 80% of individuals at some stage in their lives. Conventional diagnostic strategies—such as clinical evaluations, physical assessments, and radiological imaging (X-rays, CT scans, MRIs)—remain central to identifying spinal pathologies. However, these techniques are resource-intensive, require specialist interpretation, and often do not correlate precisely with the patient's subjective pain experience. These limitations have prompted a growing interest in data-driven and automated diagnostic systems aimed at enhancing early detection and reducing dependency on traditional workflows.

Recent progress in Machine Learning (ML) has enabled the extraction of actionable insights from medical datasets, supporting classification tasks in various domains of spinal health. Prior research has explored the application of supervised learning algorithms for detecting structural spinal abnormalities. For instance, Suri et al. (2012) [8] analyzed radiographic indicators such as disc space narrowing and vertebral degeneration. Similarly, Koivisto et al. (2014) [9] employed logistic regression and decision trees to classify lumbar pathologies based on clinical and imaging features. Yet, many of these approaches rely on limited feature sets or predefined labels, often suffering from overfitting or poor generalizability to new datasets.

A significant challenge in LBP classification using machine learning is the inherent class imbalance in medical datasets, where instances of abnormal spinal conditions are often far fewer than normal cases. This imbalance can lead to biased models with poor performance in identifying minority classes. The Synthetic Minority Over-Sampling Technique (SMOTE), introduced by Chawla et al. [10], has been widely used to address this issue by generating synthetic minority class samples. Recent study [11] have shown that SMOTE, when combined with appropriate classifiers, can significantly improve the detection of minority conditions in healthcare. However, the effectiveness of SMOTE can be limited by the complexity of the data and the heterogeneity of clinical presentations [12].

While traditional machine learning algorithms like Decision Trees (DT) and Support Vector Machines (SVM) have shown promise in LBP classification, Artificial Neural Networks (ANNs) have demonstrated superior performance in certain contexts. For instance, Richard Kijowski et al. [13] highlighted the effectiveness of deep learning models for diagnosing musculoskeletal conditions using medical imaging like radiographs, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and nuclear medicine.

To address these limitations, the present study proposes a structured and data-driven methodology for LBP classification using machine learning. The primary objective is to develop a comprehensive ML pipeline that preprocesses an International standard dataset on LBP. This includes outlier removal to improve data quality, histogram-based feature visualization to understand variable distributions, and application of SMOTE to correct class imbalance—each step intended to enhance downstream model performance.

In the subsequent phase, the study systematically evaluates and compares the performance of three widely-used ML algorithms-Decision Tree, Support Vector Machine, and Artificial Neural Network on both the original imbalanced dataset and the SMOTE-balanced version. By analysing performance metrics such as accuracy, precision, recall, and F1-score, the aim is to identify the most effective model for automatic LBP classification. This approach not only bridges gaps in earlier works that isolated preprocessing from classification but also sets the groundwork for scalable, generalizable diagnostic models applicable to broader musculoskeletal conditions.

3. Materials and Methods

The automatic prediction of LBP is important to reduce an individual's pain. To achieve this, an ML model using the International standard database is proposed. The data are collected and analysed, which involves the following processes:

- The types of features and targets in the database are identified and processed based on need.
- The features are explored using a histogram plot to identify their importance.

- The outliers are identified and removed using the trimming method.
- Imbalanced data are converted to balanced data using the SMOTE technique.

Next to data analysis, the data are partitioned and passed through the ML model, such as DT, SVM, and ANN, for training and testing. The outcome of the ML model during the testing phase is evaluated using the metrics. The metrics are used to finalize the best model for LBP prediction.

3.1. Exploratory Data

While LBP is prevalent, the intensity and impact of its symptoms can vary widely from person to person. In contrast to the mild, intermittent pain that a degenerating disc may cause, a simple lower back muscle strain may be severe enough to need an urgent care visit. The Kaggle dataset on LBP [14] was the source of the data. This dataset aims to determine whether a given individual's spinal structure contains LBP. The data include both normal and LBP records. The number of records in normal and LBP is 100 and 210.

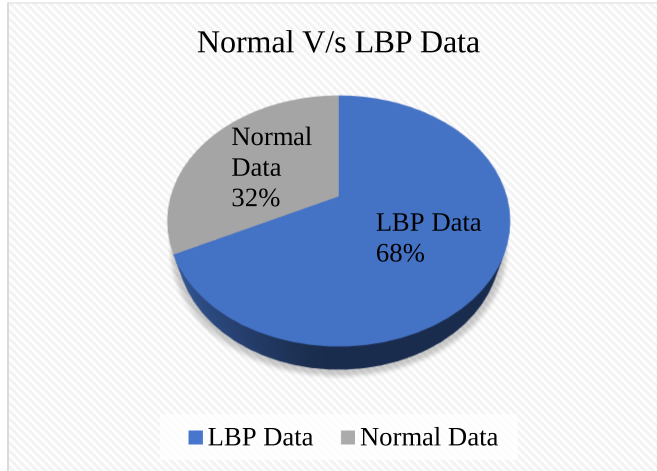


Fig. 1 Normal vs LBP data

The distribution of LBP versus normal data is given in the pie chart as shown in Figure 1. The sample data from the dataset is depicted in Table 2. The first 12 rows in the table show the features, while the last row represents the target for various subjects.

Table 2. Sample LBP data

Features	**Subject 1**	**Subject 2**	**Subject 3**	**Subject 4**	**Subject 5**
Pelvic Incidence	63.03	39.06	68.83	69.30	49.71
Pelvic Tilt	22.55	10.06	22.22	24.65	9.65
Lumbar Lordosis Angle	39.61	25.02	50.09	44.31	28.32
Sacral Slope	40.48	29.00	46.61	44.64	40.06
Pelvic Radius	98.67	114.41	105.99	101.87	108.17
Degree Spondylolisthesis	-0.25	4.56	-3.53	11.21	7.92
Pelvic Slope	0.74	0.42	0.47	0.37	0.54
Direct Tilt	12.57	12.89	26.83	23.56	35.49
Thoracic Slope	14.54	17.53	17.49	12.71	15.95
Cervical Tilt	15.30	16.78	16.66	11.42	8.87
Sacrum Angle	-28.66	-25.53	-29.03	-30.47	-16.38
Scoliosis Slope	43.51	16.11	19.22	18.83	24.92
Class	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal

3.1.1. Data Type Analysis

The data type is analyzed in this section using Table 3. The data holds 13 attributes, out of which 12 are features, and the remaining 1 is a target variable. All features are in float type, and the target variable is in categorical type. The categorical data are not suited for constructing the ML model. Using the Label encoder approach, the target variable is converted to numbers. The "0" is replaced in the place of "Abnormal", and the "1" is replaced in the place of "Normal".

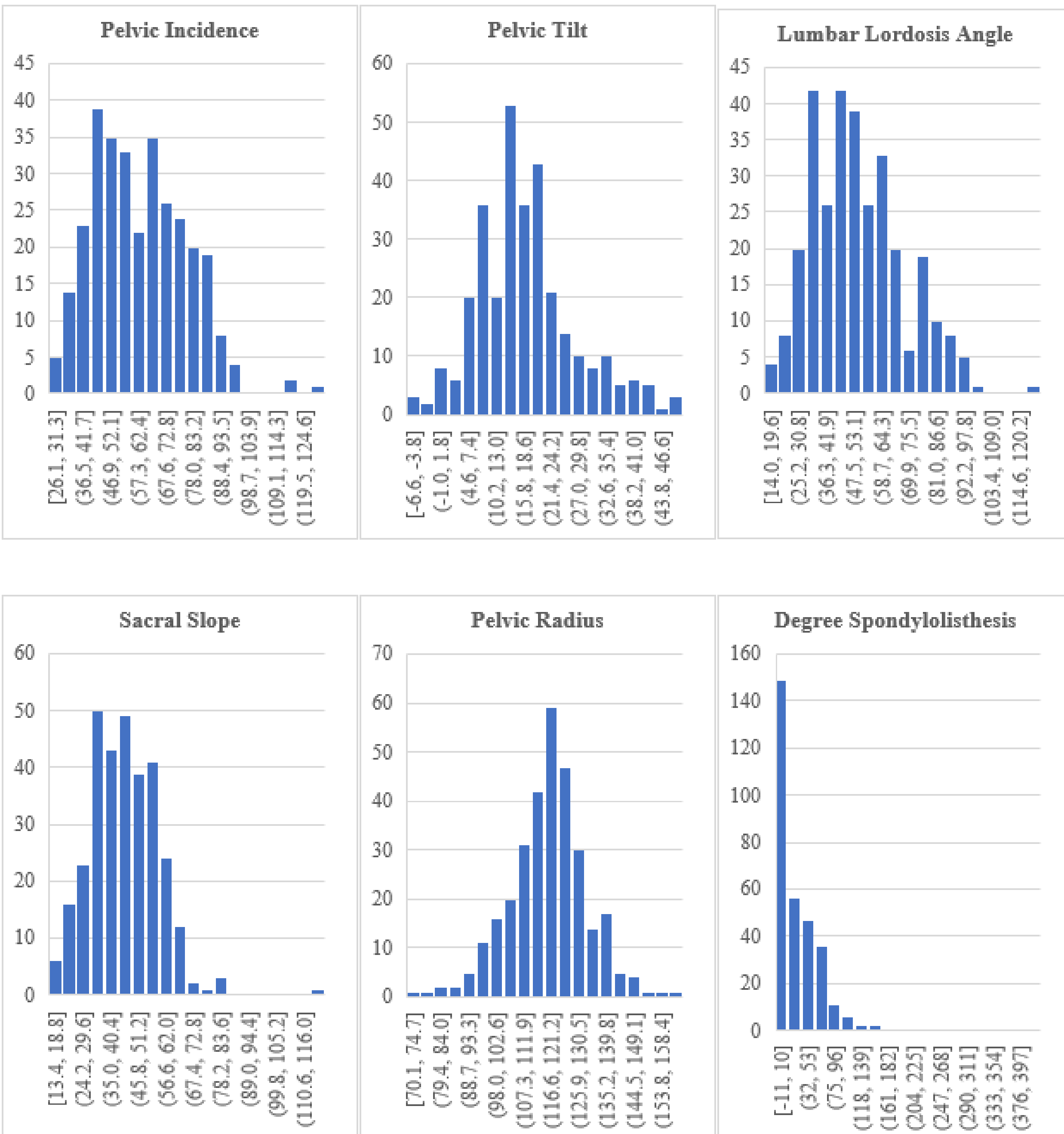
Table 3. LBP features and target data types

Attribute name	**Type**
Pelvic Incidence	numeric, float64
Pelvic Tilt	numeric, float64
Lumbar Lordosis Angle	numeric, float64
Sacral Slope	numeric, float64
Pelvic Radius	numeric, float64
Degree Spondylolisthesis	numeric, float64
Pelvic Slope	numeric, float64
Direct Tilt	numeric, float64
Thoracic Slope	numeric, float64
Cervical Tilt	numeric, float64
Sacrum Angle	numeric, float64
Scoliosis Slope	numeric, float64
Attribute Class	categorical, object

3.1.2. Histogram of Each Feature

The ML model's features are important because they determine the accuracy and execution time. It is very important to identify and process the most correlated features. The histogram is employed to learn more details about the features. Histograms are used to show the distribution of numerical data graphically. It is said to have been invented by Karl Pearson. A histogram representation is created in two stages [15].

- The first stage is to establish a grid of discrete value ranges. Each interval comprises non-overlapping, consecutive, adjacent, and equally sized bins.
- The frequency of each interval, which is just the total number of values within that interval, is determined in the second stage. Figure 2 depicts the histogram plot for 12 features.



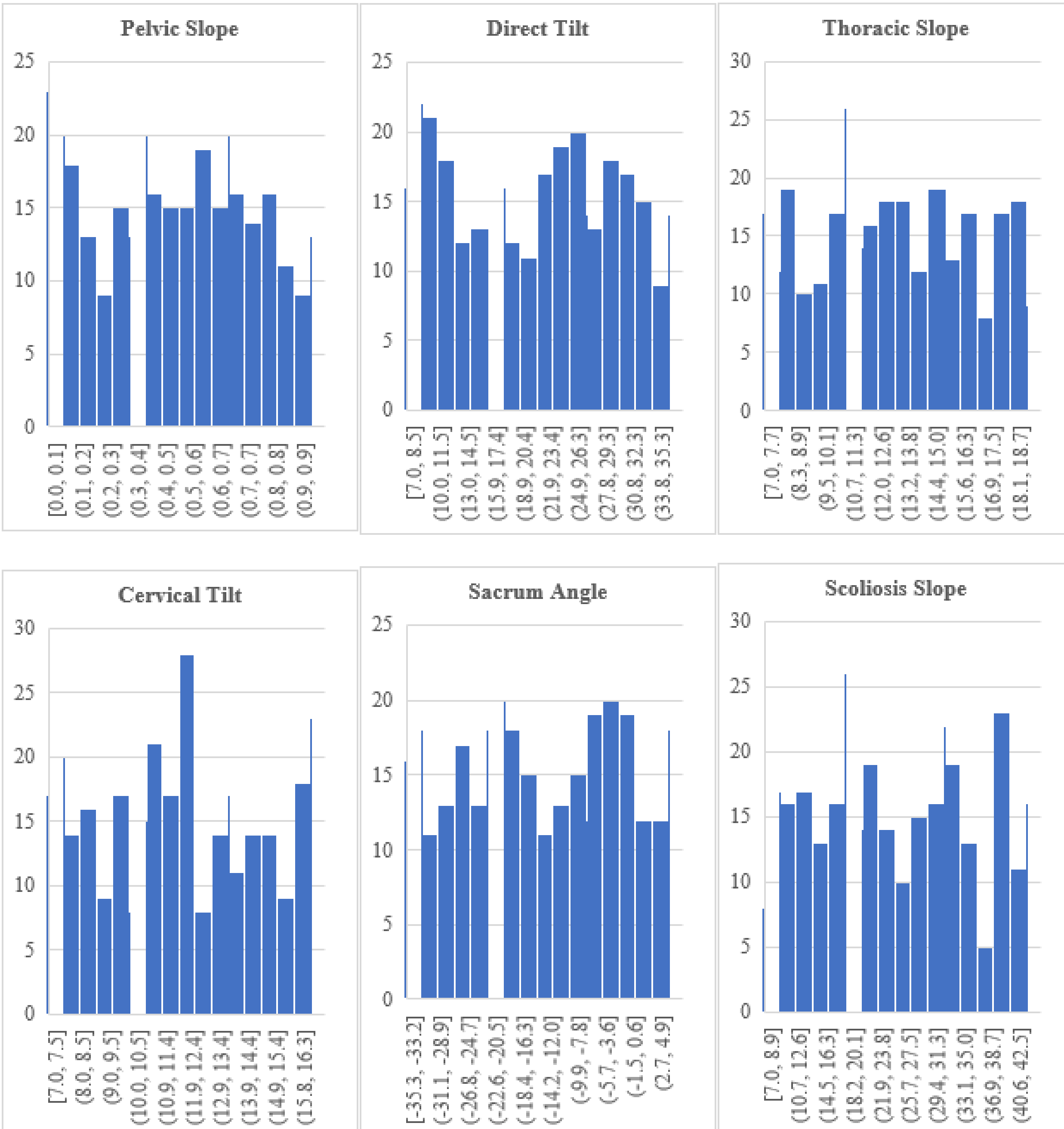
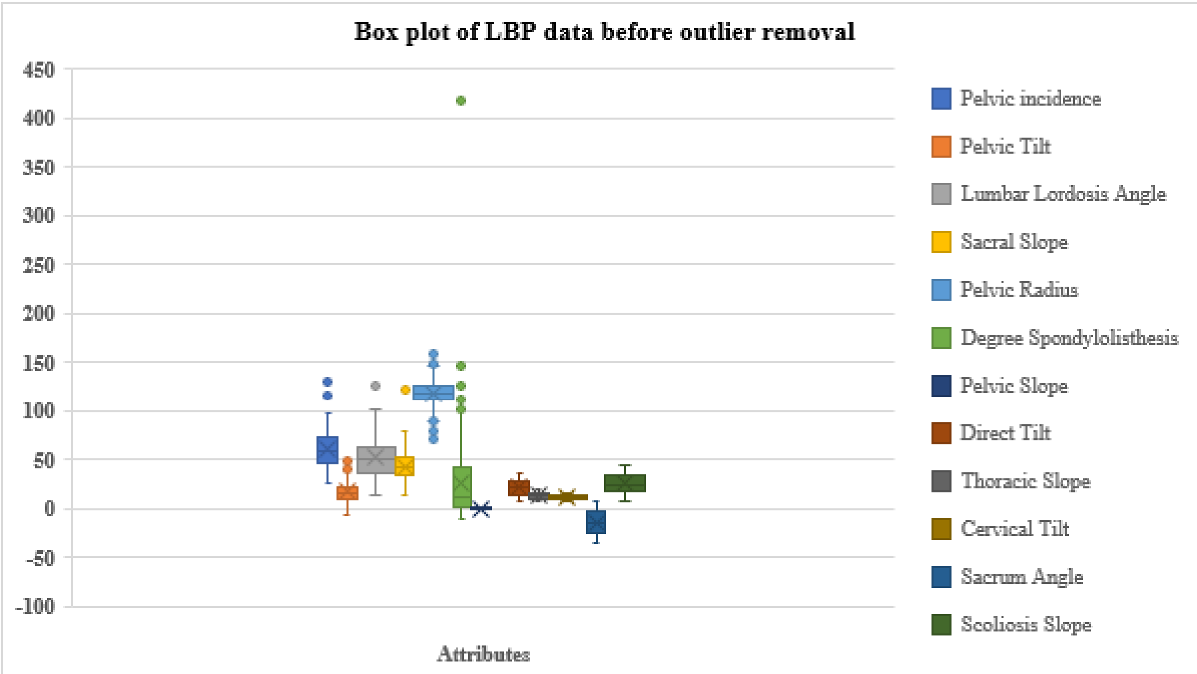


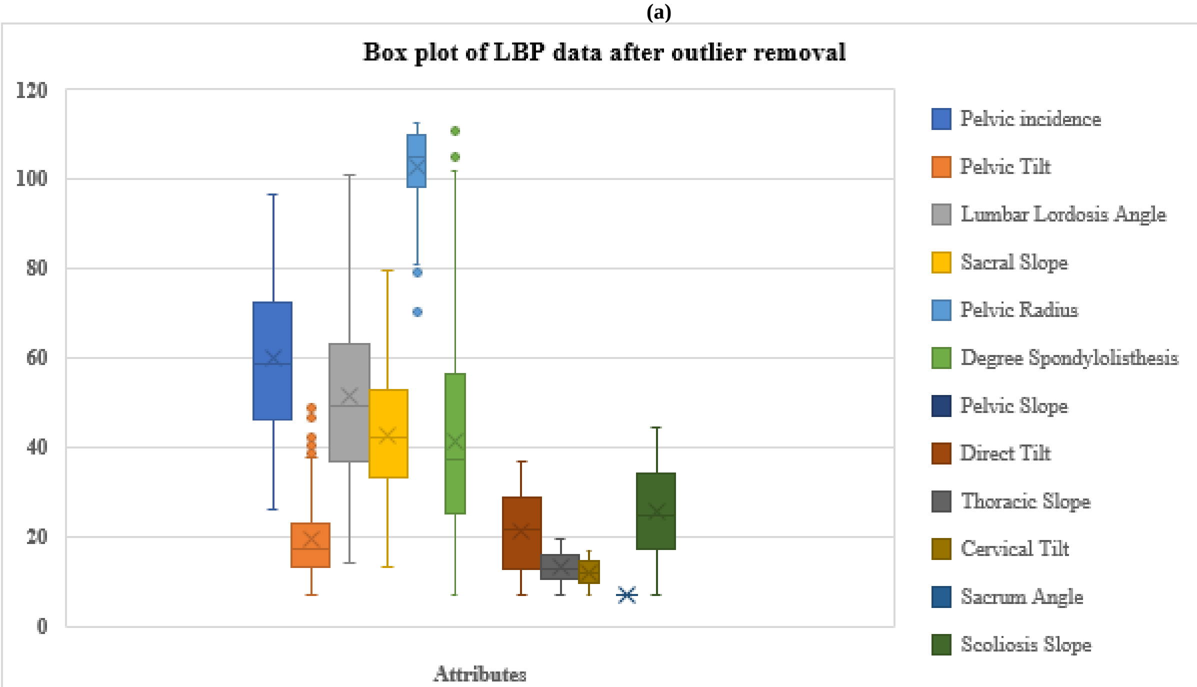
Fig. 2 Histogram plot of features

3.1.3. Outlier Removal

A value deviating significantly from the acquired data is called an outlier. The highest and lowest points in a data set are known as outliers. The Box and Whisker Graph is a simple way to visualize outliers [16]. A Box Plot is a data visualization in which information is presented using boxy shapes. The graph divides enormous, complex data sets into understandable quartiles and means. This image will help you locate any unusual outliers in your LBP data. The Box Plot, as shown in Figure 3(a), categorizes key indicators into four equal groups or quartiles. An outlier is a number outside the data's typical range. In other words, it is a number that does not fit the normal distribution and may affect the entire collection of numbers. Outlier values are considered outliers that may bias the conclusions. An outlier is a statistic significantly higher or smaller than the median observation by more than 1.5, according to experts in data visualization.



(a)



(b)

Fig. 3 Box plot of LBP data before and after outlier removal

The box represents the center half of the data, while the line in the middle represents the median. The lines extending from the box [17] show the range of the remaining data. A data point that deviates significantly from the mean or median is referred to as an outlier. An outlier is a red dot in Figure 3(a) that does not fit into the boxplot. To reduce the outliers in the LBP dataset, "trimming" is employed to delete all undesirable data. The box plot of the LBP data after outlier elimination is shown in Figure 3(b).

3.1.4. Data Balancing

The acquired data has a class imbalance, meaning that the total number of instances in each class is not equal. The challenges of learning from class-imbalanced data have recently gained traction in various disciplines [18]. Evaluation of classifier performance is essential in this setting due to the class imbalance's significant effects on the learning strategy, which typically produces classification algorithms that have poor predicted accuracy for the minority class and a bias towards categorizing most new samples into the majority class.

The SMOTE [19] method is an oversampling technique that artificially generates samples from minority populations. It is widely used because it has the potential to surpass traditional oversampling. To train the classifier, a synthetic training set balanced across classes or close to it must be obtained. Using the equation below, SMOTE samples are created by combining two minority-group samples (x and x_R).

$$s = x + u \cdot (x_R - x), 0 \leq u \leq 1 \qquad (1)$$

Where,
s: The synthetic sample.
x: A randomly selected sample from the minority class.
x_R: A nearest neighbor of x, randomly selected from the minority class.
u: A random number drawn uniformly from the range [0,1].

The closest member of the minority class to x is selected randomly as xR. Table 4 shows the count of data used in each stage. The number of abnormal and normal entries in raw data is 210 and 100. After SMOTE, the count changes to 210 and 210. Of the 420 samples, 336 were employed for training, and 84 were used for testing.

Table 4. Data count in each stage

LBB Data	Raw Data	SMOTE Data	Train Data	Test Data
Abnormal	210	210	168	42
Normal	100	210	168	42

3.2. Machine Learning Model

The ML models, such as Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN), are used to identify normal and normal LBP data. The workings of those ML models are discussed in this section.

3.2.1. Decision Tree

DT is a successful method frequently employed across many fields [20]. In each DT, a numerical characteristic is compared to a threshold value, and the model efficiently and consistently builds on the preceding one. Unlike the numerical weights employed in NN inter-node connections, conceptual rules are much easier to construct [21]. Every tree has a trunk, branches, and leaves. Each node in the target classification represents a set of features, and a set of subsets defines its value. DT has gained popularity due to its ease of use and accuracy across various data types. Entropy can be used to determine the purity or randomness of a dataset. Entropy levels frequently lie between zero and one. The closer its value is to 0, the better it is. Its value is worse when it is equal to 1. The entropy of classifying set S concerning c states by solving equation (1) if the target is G with variable attribute values may be calculated.

$$\boldsymbol{Entropy(S)} = \sum_{i=1}^{c} -\boldsymbol{p_i \log_2 p_i} \qquad \mathbf{(2)}$$

Where,
P_i → ratio of the number of samples in the group to the value of the i^{th} feature.

Information gain, or mutual information, is a measure used in the segmentation process. This provides a natural measure of how much insight one has into the value of a random variable.

In contrast to entropy, its value should increase over time. Benefits from data collection based on the concept of entropy in Equation (2), $Ga(S, A)$ may be written as follows:

$$Gain(S, A) = E(S) - \sum_{v \epsilon V(A)} \frac{|S_v|}{|S|} E(S_v) \qquad (3)$$

Where,
S: The set of all samples in the dataset.
A: The attribute for which information gain is being calculated.
$V(A)$: The set of all possible values (range) of attribute A.
S_v: The subset of SSS containing only those samples for which attribute A has value v.
$|S|$: The total number of samples in the dataset S.
$|S_v|$: The number of samples in the subset S_v.
$E(S)$: The entropy of the dataset S.
$E(S_v)$: The entropy of the subset S_v

The flow model for the Decision Tree can be visualized in Figure 4. The selection of appropriate hyperparameters is known to be crucial for achieving optimal performance in decision tree models. In this study, the hyperparameters for the Decision Tree model were tuned through a grid search approach combined with stratified k-fold cross-validation (k=5).

Various values for the maximum depth of the tree, the minimum number of samples required to split an internal node, and the minimum number of samples required to be at a leaf 1 node were explored.

The optimal hyperparameter set was selected based on the highest average F1-score that was achieved across the validation folds during the cross-validation process. This tuning process was performed with the aim of optimizing the generalization performance of the decision tree on the LBP classification task.

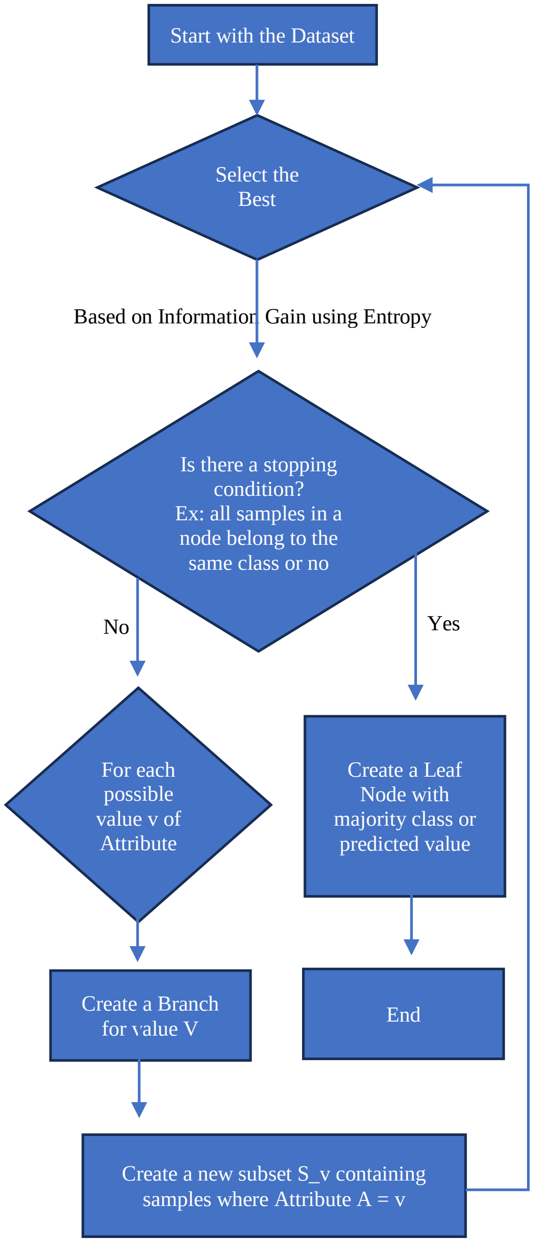


Fig. 4 Decision tree flow model

3.2.2. Support Vector Machine

SVM is a popular ML technique for handling classification tasks. It was created in the early 1990s by Vladimir Vapnik and his colleagues [22]. Finding the ideal choice boundary for categorizing the data point is required to separate the classes in an n-dimensional space from the various decision lines/boundaries so that the new data point can be easily classified. The SVM hyperplane represents the ideal border. Every hyperplane must include a maximum margin representing the highest feasible separation between data points. SVM selects the points and vectors at the extremes to form the hyperplane. The data points or vectors nearest to the hyperplane influence its position. A support vector, as the name implies, supports a hyperplane. SVM comes in two types; the non-linear version is used. Non-linear SVM is employed when the data is not neatly separated. The non-SVM classifier on LBP datasets is used for prediction. When applied to a dataset, a kernel function transforms a non-linear decision surface into a linear equation in higher-dimensional space. Figure 5 visualizes the mechanism by which SVM constructs a separating hyperplane. Instead of simply drawing any boundary, SVM identifies specific data points-termed support vectors-that are most influential in defining the optimal separation. These vectors lie closest to the decision surface and are instrumental in determining its orientation and placement. By focusing on these critical points, SVM ensures that the margin between classes is maximized, leading to a more reliable and generalized classifier capable of handling complex classification tasks.

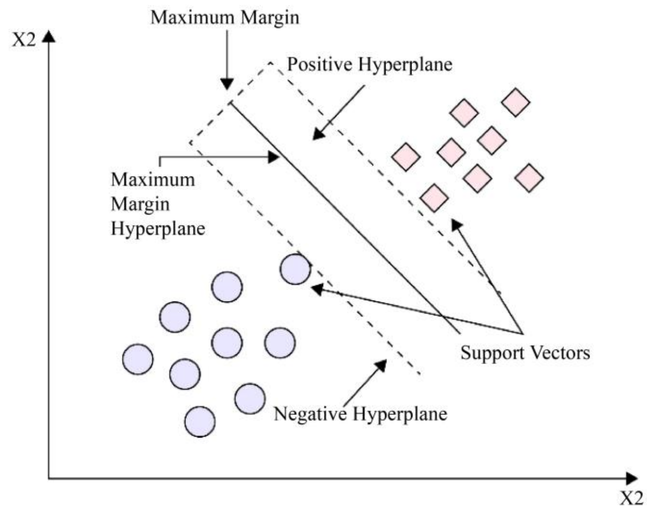


Fig. 5 Hyperplane of a nonlinear SVM

The performance of Support Vector Machine models is understood to be significantly influenced by the choice of hyperparameters. In this study, the hyperparameters for the Support Vector Machine model were tuned using a grid search approach combined with stratified k-fold cross-validation (k=5). Different kernel functions (linear, radial basis function - RBF, and polynomial), as well as the regularization parameter (C) and the kernel coefficient (gamma for RBF and polynomial kernels), were experimented with. The optimal hyperparameter set was selected based on the highest average F1-score that was achieved across the validation folds during the cross-validation process. This tuning was conducted with the goal of optimizing the generalization capability of the support vector machine for the LBP classification task.

3.2.3. Artificial Neural Network

The biological NN of the human brain served as inspiration for the popular ML technique known as ANN [23]. In Feed-Forward Neural Networks (FFNN) [24], the weight parameter of each neuron is passed as output to the subsequent layer. A major subgroup of FFNN is the Multilayer Perceptron

(MLP) [25]. The backpropagation technique is the most commonly used approach for training MLPs. The weights between the neurons are modified to accomplish this. This model is particularly effective in learning patterns. It’s adaptable enough to handle data changes, but there’s a danger that the system will converge slowly and end up at a local maximum. The big challenge is how many layers a network should have, how many neurons should be in the hidden layer, and what connections should be created between them. These characteristics have a significant impact on the effectiveness of the ANN. Any of these measurements could show considerable variation. The results of applying different ANN architectures to a problem space will differ. However, trial and error are required to arrive at the best potential ANN architecture.

The training data sample calculated the ANN’s neuron and bias weight variables. Adjusting the number of neurons and training epochs achieved the lowest feasible error rate.

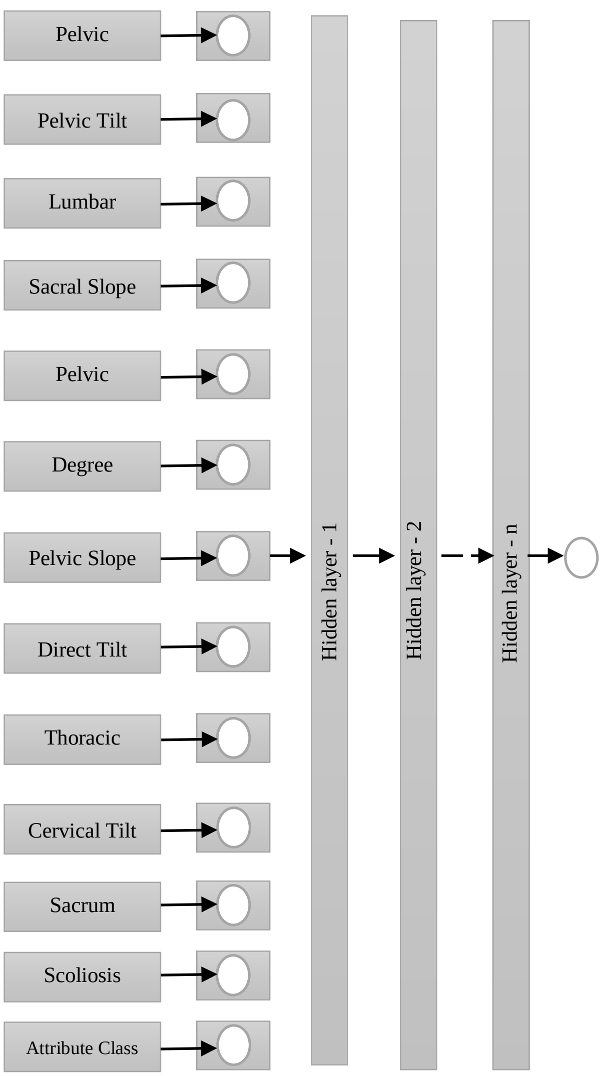


Fig. 6 ANN architecture

Later, the trained network was applied to the test data. Figure 6 shows that the implemented ANN has 12 input neurons in the input layer, six hidden layers, and one neuron in the output layer.

The architecture and training parameters, or hyperparameters, are recognized as critical factors in the performance of artificial neural networks. In this study, the hyperparameters for the Artificial Neural Network model were tuned through a grid search approach combined with stratified k-fold cross-validation (k=5). Variations in the number of hidden layers and the number of neurons in each hidden layer, the activation function (ReLU), and the learning rate of the Adam optimizer were considered. The optimal hyperparameter set was selected based on the highest average F1-score that was achieved across the validation folds during the cross-validation process. This tuning was performed to optimize the learning and generalization abilities of the neural network on the LBP classification task.

4. Result and Discussion

The outcome of the ML model on LBP identification using balanced data by SMOTE and imbalanced data are discussed in this section. The collected data should go through the processing technique of data type conversion, feature exploration, and outlier removal. The processed data without balancing is given to the three ML models for classification. The reported performance metrics—accuracy, specificity, sensitivity, precision, and F1-score—provide a comprehensive evaluation of the machine learning models’ effectiveness in classifying LBP. Accuracy indicates the overall correctness of the model’s predictions, representing the proportion of correctly classified instances out of the total. In a medical context like LBP diagnosis, high accuracy is desirable to minimize both false positives and false negatives. Specificity, or the true negative rate, is crucial as it measures the model’s ability to correctly identify individuals without LBP, thereby reducing unnecessary anxiety and further investigations for healthy individuals. Conversely, sensitivity, or the true positive rate, highlights the model’s capability to correctly identify individuals who do have LBP, ensuring that those in need of diagnosis and potential treatment are not missed. Precision focuses on the positive predictive value, indicating the proportion of correctly identified LBP cases out of all instances predicted as LBP. High precision minimizes the burden on the healthcare system from false positive diagnoses. Finally, the F1-score, being the harmonic mean of precision and sensitivity, provides a balanced measure, particularly important in datasets with potential class imbalance even after SMOTE, as it considers both false positives and false negatives. The consistently high values observed across all metrics for the ANN model on the SMOTE-balanced data suggest its strong ability to correctly classify both LBP and non-LBP cases with a low rate of errors in both directions, making it a promising tool for assisting in LBP diagnosis. The accuracy, sensitivity, specificity, precision, and F1-score performance measures are analyzed for all models. The DT gives an accuracy of 89.29%, a specificity of 92.50%, a sensitivity of 86.36%, a precision of 92.50%, and an F1-score of 89.41%. Next, the SVM outcome is analyzed, and the results of the metrics are 90.48%, 92.86%, 88.10%, 92.68%, and 90.24%. Finally, the ANN is taken, and the accuracy is 94.05%, specificity and sensitivity values are 95.24% and 92.86%, the precision score is 95.12%, and the value of F1 is 93.98%.

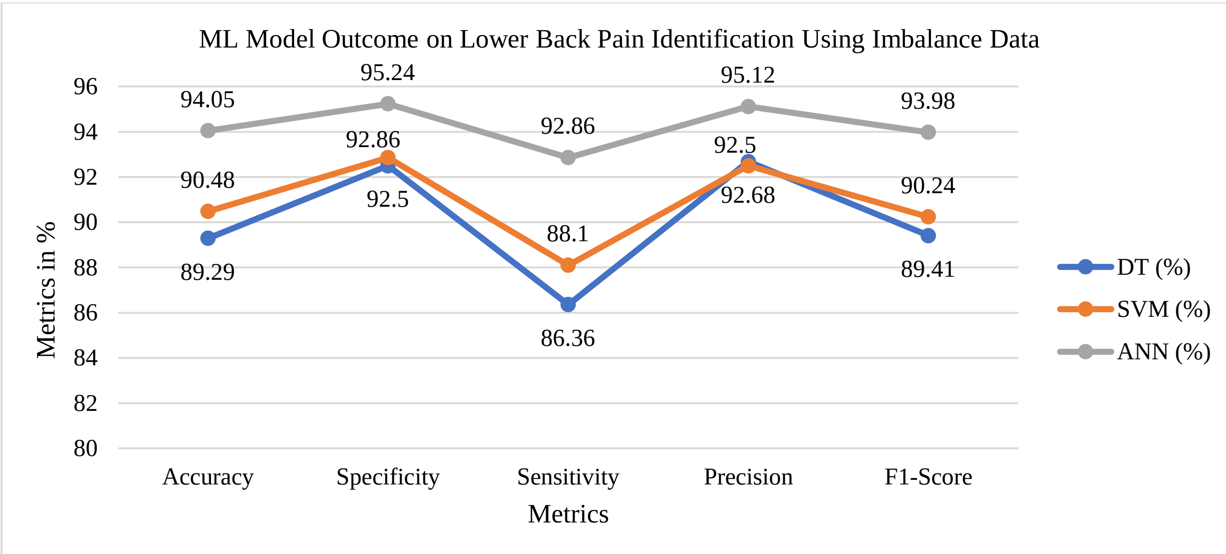


Fig. 7 Outcome of ML model on imbalanced data of LBP

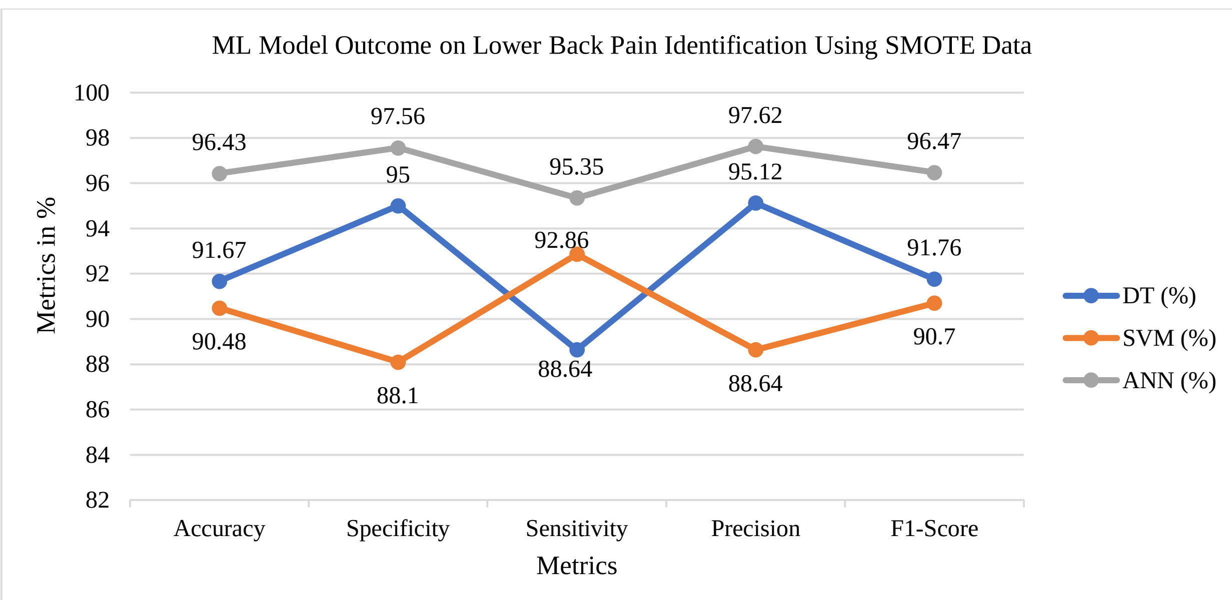


Fig. 8 Outcome of ML model on SMOTE data of LBP

Figure 7 plots the metrics score of the ML model using imbalanced data. The plot ANN gives the maximum value in all metrics.

Three ML models are used for classification based on the balanced processed data using the SMOTE approach. Again, all models' effectiveness is evaluated using the same performance measure used in the above condition. The accuracy for the DT is 90.48%, while its specificity, sensitivity, precision, and F1-score are 88.10%, 92.86%, 88.64%, and 90.70%. The next step involves analyzing the SVM results, which yielded the following metric values: 91.67%, 95%, 88.64%, 95.12%, and 91.76%. Finally, ANN is utilized, yielding v96.43% accuracy, 97.56% specificity, 95.35% sensitivity, 97.62% precision, and a 96.47% F1 value. Metrics for the ML model's performance on the SMOTE data are plotted in Figure 8. The ANN plot shows the maximum metrics score.

Confusion matrices, constructed from the original dataset of 310 subjects (210 LBP, 100 non-LBP), provide a detailed interpretation of the machine learning models' diagnostic performance by presenting True Positives (TP), true negatives (TN), False Positives (FP), and False Negatives (FN). This analysis is crucial for understanding each model's ability to accurately diagnose patients, minimizing both missed detections and false alarms.

Table 5 summarizes the confusion matrix counts for the Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN) models. The DT model correctly identified 181 LBP and 93 non-LBP cases, with 29 false negatives and 7 false positives. The SVM showed slightly improved performance with 185 true positives and 93 true negatives, maintaining 7 false positives and reducing false negatives to 25. The ANN demonstrated the strongest performance, achieving 195 true positives and 95 true negatives, with only 5 false positives and 15 false negatives, highlighting its superior learning ability with the imbalanced data.

Table 5. Confusion matrix counts for ML models

ML Model	True Positives (TP)	True Negatives (TN)	False Positives (FP)	False Negatives (FN)
DT	181	93	7	29
SVM	185	93	7	25
ANN	195	95	5	15

This refined confusion matrix analysis confirms the ANN's outperformance over DT and SVM in minimizing classification errors on the original LBP dataset. The ANN's superior performance can be attributed to its capacity to learn complex non-linear relationships and hierarchical feature representations from the 12 input features, which is advantageous given the potentially intricate interplay among these features in predicting LBP. In contrast, the DT's hierarchical structure might not efficiently capture subtle patterns, and the SVM's performance is highly dependent on hyperparameter tuning and kernel choice, potentially limiting its ability to model highly non-linear decision boundaries. The ANN's ability to learn complex patterns directly from data, coupled with the benefits of data balancing techniques like SMOTE (as indicated by the analysis of Figures 7 and 8 showing improved results with balanced data), likely enabled it to achieve better discrimination between LBP and non-LBP cases compared to the DT and SVM models. Comparing the performance metrics of models trained on original versus SMOTE data reveals the importance of data balancing for achieving better results, and among the models trained on SMOTE data, ANN consistently yields the best performance.

5. Conclusion

Low Back Pain (LBP) remains the leading cause of disability and lost productivity worldwide, and while treatment options have advanced, achieving accurate, consistent diagnosis continues to challenge clinical workflows. Traditional diagnostic approaches, though well-established, often fall short due to their reliance on subjective interpretation, high resource demand, and limited ability to correlate with patient-reported symptoms. In contrast to earlier studies that either employed limited Machine Learning (ML) techniques or neglected the role of data quality and balance, this study presents a focused and methodical solution to the LBP classification problem using ML.

Utilizing the LBP dataset comprising 12 input features and a binary classification target, this work addressed one of the most critical challenges in medical data analysis: class imbalance. With 210 "abnormal" cases and only 100 "normal" ones, the dataset risked biasing any model trained without proper rebalancing. To resolve this, the Synthetic Minority Over-sampling Technique (SMOTE) was employed, effectively equalizing class distributions and enhancing model fairness.

A comprehensive pipeline was implemented to evaluate ML performance both before and after SMOTE integration. Three models—Decision Tree (DT), Support Vector Machine (SVM), and Artificial Neural Network (ANN)—were assessed on key performance metrics and visualized using line plots for comparative clarity. The results unequivocally confirmed the critical impact of SMOTE, with all models showing improved accuracy, precision, and recall on the balanced dataset. Notably, the ANN model consistently outperformed DT and SVM, demonstrating superior classification accuracy and minimal error.

Compared to prior studies that either overlooked class imbalance or relied on conventional ML techniques, the proposed SMOTE-enhanced approach provides a more robust and fair classification framework. For instance, Sadeghi et al. [26] developed ML models to predict chronic LBP risk using national health survey data but did not address class imbalance, potentially limiting model generalizability. Similarly, Abujaber et al. [27] utilized surface electromyography (sEMG) signals for LBP classification without implementing data balancing techniques, which could affect classification fairness. In contrast, this study not only corrects class skew using SMOTE but also systematically demonstrates its influence on classification metrics. The significant performance gain observed, particularly with the ANN model, substantiates the value of integrating data balancing techniques into diagnostic pipelines for LBP and related conditions. These results reinforce the argument that advanced preprocessing, when coupled with deep learning, offers a clinically viable and scalable solution for musculoskeletal disorder detection. This study not only validates the importance of preprocessing strategies such as outlier removal and SMOTE but also establishes ANN as the most effective classifier for automatic LBP detection in the tested scenario. By integrating balanced data handling with robust model evaluation, the findings offer a decisive advancement over previous fragmented approaches and provide a scalable framework for future diagnostic systems in musculoskeletal health.

References

[1] D. Hoy et al., "The Global Burden of Low Back Pain: Estimates from the Global Burden of Disease 2010 Study," *Annals of the Rheumatic Diseases*, vol. 73, no. 6, pp. 968-974, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[2] *Health, United States*, National Center for Health Statistics, U.S. Department of Health and Human Services, 2020. [Google Scholar] [Publisher Link]

[3] Gary S. Firestein, and Gary Koretzky, *Firestein & Kelley's Textbook of Rheumatology*, 12^{th} ed., Elsevier, pp. 740-763, 2024. [Google Scholar] [Publisher Link]

[4] G.B. Andersson, "Epidemiological Features of Chronic Low-Back Pain," *The Lancet*, vol. 354, no. 9178, pp. 581-585, 1999. [Google Scholar] [Publisher Link]

[5] Michael T. Modic, and Jeffrey S. Ross, "Lumbar Degenerative Disk Disease," *Radiology*, vol. 245, no. 1, pp. 43-61, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[6] Amir Jamaludin, Timor Kadir, and Andrew Zisserman, "SpineNet: Automated Classification and Evidence Visualization in Spinal MRIs," *Medical Image Analysis*, vol. 41, pp. 63-73, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Joachim Oertel et al., "Acute Low Back Pain: Epidemiology, Etiology, and Prevention: WFNS Spine Committee Recommendations," *World Neurosurgery: X*, vol. 22, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Pradeep Suri et al., "Does Lumbar Spinal Degeneration begin with the Anterior Structures? A Study of the Observed Epidemiology in a Community-Based Population," *BMC Musculoskeletal Disorders*, vol. 12, pp. 1-7, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[9] Katri Koivisto et al., "Efficacy of Zoledronic Acid for Chronic Low Back Pain Associated with Modic Changes in Magnetic Resonance Imaging," *BMC Musculoskeletal Disorders*, vol. 15, no. 1, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[10] N.V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[11] Surani Matharaarachchi, Mike Domaratzki, and Saman Muthukumarana, "Enhancing SMOTE for Imbalanced Data with Abnormal Minority Instances," *Machine Learning with Applications*, vol. 18, pp. 1-31, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Alberto Fernandez et al., "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863-905, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] Richard Kijowski et al., "Deep Learning for Lesion Detection, Progression, and Prediction of Musculoskeletal Disease," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 6, pp. 1607-1619, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Lower Back Pain Symptoms Dataset, Kaggle. [Online]. Available: https://www.kaggle.com/datasets/sammy123/lower-back-pain-symptoms-dataset?datasetId=107&sortBy=voteCount

[15] Shashwati Mishra, and Mrutyunjaya Panda, "A Histogram-Based Classification of Image Database Using Scale-Invariant Features," *International Journal of Image, Graphics and Signal Processing*, vol. 9, no. 6, pp. 55-64, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] K. Andrea, G.L. Shevlyakov, and P.O. Smirnov, "*Detection of Outliers with Boxplots*," Belarusian State University, Minsk, Belarus, Thesis, pp. 141-144, 2013. [Google Scholar] [Publisher Link]

[17] David F. Williamson, Robert A. Parker, and Juliette S. Kendrick, "The Box Plot: A Simple Visual Method to Interpret Data," *Annals of Internal Medicine*, vol. 110, no. 11, pp. 916-921, 1989. [CrossRef] [Google Scholar] [Publisher Link]

[18] H. He, and E.A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[19] Rok Blagus, and Lara Lusa, "SMOTE for High-Dimensional Class-Imbalanced Data," *BMC Bioinformatics*, vol. 14, pp. 1-16, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[20] Gary Stein et al., "Decision Tree Classifier for Network Intrusion Detection with GA-based Feature Selection," *Proceedings of the 43^{rd} Annual Southeast Regional Conference*, vol. 2, pp. 136-141, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[21] P.H. Swain, and H. Hauska, "The Decision Tree Classifier: Design and Potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142-147, 1977. [CrossRef] [Google Scholar] [Publisher Link]

[22] Corinna Cortes, and Vladimir Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, 1995. [CrossRef] [Google Scholar] [Publisher Link]

[23] J.J. Hopfield, "Artificial Neural Networks," *IEEE Circuits and Devices Magazine*, vol. 4, no. 5, pp. 3-10, 1988. [CrossRef] [Google Scholar] [Publisher Link]
[24] G. Bebis, and M. Georgiopoulos, "Feed-Forward Neural Networks," *IEEE Potentials*, vol. 13, no. 4, pp. 27-31, 1994. [CrossRef] [Google Scholar] [Publisher Link]
[25] Kit Yan Chan et al., "Deep Neural Networks in the Cloud: Review, Applications, Challenges and Research Directions," *Neurocomputing*, vol. 545, pp. 1-24, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[26] Jae-Geum Shim et al., "Machine Learning Approaches to Predict Chronic Lower Back Pain in People Aged Over 50 Years," *Medicina*, vol. 57, no. 11, pp. 1-9, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[27] Bernard X.W. Liew et al., "Interpretable Machine Learning Models for Classifying Low Back Pain Status Using Electromyographic and Kinematic Data," *European Spine Journal*, vol. 29, pp. 1845-1859, 2020. [CrossRef] [Google Scholar] [Publisher Link]